CHAPTER

Schistosome Long Terminal Repeat Retrotransposons

Claudia S. Copeland, Thewarach Laha and Paul J. Brindley*

Abstract

The human schistosomes, blood flukes of the Genus *Schistosoma*, have a large genome estimated to be at least 270 megabase pairs (haploid) in size, arrayed on eight pairs (2n = 16) of chromosomes including the Z (male) and W (female) sex chromosomes. The genome appears to include about 14,000 protein encoding genes and a large amount of repetitive sequences. Much of this repetitive component of the schistosome genome is comprised of mobile genetic elements. Of these elements, an expanding number of discrete long terminal repeat (LTR) retrotransposons have been identified in the past five years. The identity, structure, phylogenetic relationships, and size contributions of these schistosome LTR retrotransposons are reviewed here. These elements include the *gypsy*-like or *Pao/BEL* -like retrotransposons *Boudicca*, *Sinbad*, *fugitive*, *Saci-1, Saci-2, and Saci-3* from *Schistosoma mansoni* and *Gulliver* from *Schistosoma japonicum*.

Introduction

Schistosomes and Schistosomiasis

Schistosomiasis is considered the most important of the human helminthiases, and second only to malaria among tropical parasitic diseases, in terms of morbidity.[1,2] Up to 200 million people are estimated to be infected worldwide, with annual deaths as high as 200,000.[2] Symptoms range from chronic pain due to granulomatous lesions of the liver and intestines to hepatosplenomegaly, portal hypertension, ascites, diarrhea, and impaired physical and cognitive development in children.[3,4] Female genital schistosomiasis, a common condition among women infected with *Schistosoma haematobium*, can lead to abortion or premature delivery as well as lesions throughout the reproductive organs, and can facilitate sexually transmitted disease infection.[5] Moreover, interactions with other infectious diseases can induce increased pathology, as with coinfection with hepatitis C, in which liver damage can be more severe than in patients with either disease alone[6] and with HIV, where viral load has been found to be significantly higher in patients coinfected with helminth parasites including schistosomes,[5] presumably due to rapid initial viral replication facilitated by the immune response to the parasites. Furthermore, schistosomiasis is associated with bladder[7] and other cancers.[8]

The life cycle of *Schistosoma* involves parasitism of both vertebrates and snail species specific for each species of schistosome. Infectious larvae known as cercariae emerge from the snails into a body of water, where they initiate infection by direct penetration of human skin. In the

*Corresponding Author: Paul J. Brindley—Department of Tropical Medicine, SL-17, Tulane University, Health Sciences Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112, U.S.A. Email: paul.brindley@tulane.edu

Mobile Genetic Elements in Metazoan Parasites, edited by Paul J. Brindley.

human host, the worms develop into male and female adults which live together within the mesenteric venules of the intestines, and release eggs into the blood stream. To perpetuate the life cycle, the eggs exit from the blood vessels and traverse the intestinal wall, facilitated by secreted proteolytic enzymes and their spines, and pass out in the feces to fresh water. Though chemotherapy is available, its effectiveness is limited by continuous reinfection upon subsequent exposure to water containing cercariae. Furthermore, symptoms do not necessarily resolve upon chemotherapeutic cure of the infection, and chronic symptoms of the disease can remain with the patient for life. No vaccine is currently available.

Health education in combination with drug therapy forms the cornerstone of the World Health Organization's strategy to combat schistosomiasis. This strategy has met with mixed success. Though the endemic distribution of schistosomiasis has changed in the past 50 years, overall, the estimated number of infected persons and those at risk of infection has not been reduced.[9-11] It is hoped that an enhanced understanding of the schistosome genome can be expected to lead to long-term strategies for the control of schistosomiasis.

The Schistosome Genome

Schistosome species infecting humans include *Schistosoma mansoni*, *Schistosoma japonicum*, *Schistosoma haematobium*, and, to a lesser extent, *Schistosoma intercalatum* and *Schistosoma mekongi.*[12] Of these, the African blood fluke, *S. mansoni*, and the Asian blood fluke, *S. japonicum* have been characterized reasonably well at the genomic level. Schistosomes have comparatively large genomes, estimated at ~270 megabase pairs for the haploid genome of *Schistosoma mansoni*, arrayed on seven pairs of autosomes and one pair of sex chromosomes.[13] For comparison, the schistosome genome is about the same size as that of the puffer fish, *Takifugu rubripes*, two to three times the size of that of the angiosperm, *Arabidopsis thaliana*, or the free-living nematode, *Caenorhabditis elegans*, ten times the size of the *Plasmodium falciparum* genome, and about one tenth the size of the human genome. The other major schistosome species parasitizing humans probably have a genome of similar size to that of *S. mansoni*, based on the similarity in appearance of their karyotypes,[14] though very recent estimates based on genome sequencing indicate that the genome of *S. japonicum* may be as large as 400 megabase pairs or more (Wang S-Y, personal communication). Although no schistosome genome has been sequenced in its entirety, several hundred thousand schistosome expressed sequence tags (ESTs) and genome survey sequences have been deposited in GenBank, and these may represent the complete transcriptomes of *S. mansoni* and *S. japonicum*, covering ~14,000 genes in *S. mansoni* and ~13,000 genes in *S. japonicum.*[15] These ESTs appear to represent most, if not all, *S. mansoni* and *S. haematobium* genes, with expressed gene content of schistosomes previously estimated to be 15,000-20,000.[17] The mobile genetic elements (MGEs) of the schistosome genome include transposons, which replicate via a "cut and paste" process of direct transposition as DNA, retrotransposons, which replicate through an RNA mediated process, and SINE-like elements, which use the enzymes of retrotransposons to replicate (Fig. 1). Recent findings suggest that up to half of the entire schistosome genome may be comprised of repetitive sequences, with much of this repetitive complement comprised of mobile genetic elements.[18] Mobile genetic elements appear to make up at least one quarter of the schistosome genome. Here we review the LTR retrotransposons from the genomes of schistosome species. The identity, structure, phylogenetic relationships, and contribution of these elements to genome size in schistosomes are described, and we address their probable role in schistosome evolution and potential utility in applications such as introducing transgenes into schistosomes.

LTR Retrotransposons

The retrotransposon component of the schistosome genome includes both the retroviral-like LTR retrotransposons, so named because they are framed by long terminal direct repeats, and nonLTR retrotransposons, which often include A-rich 3' termini in their genomic forms. LTR retrotransposons resemble retroviruses in their structure and intracellular life cycles (Figs. 2, 3).

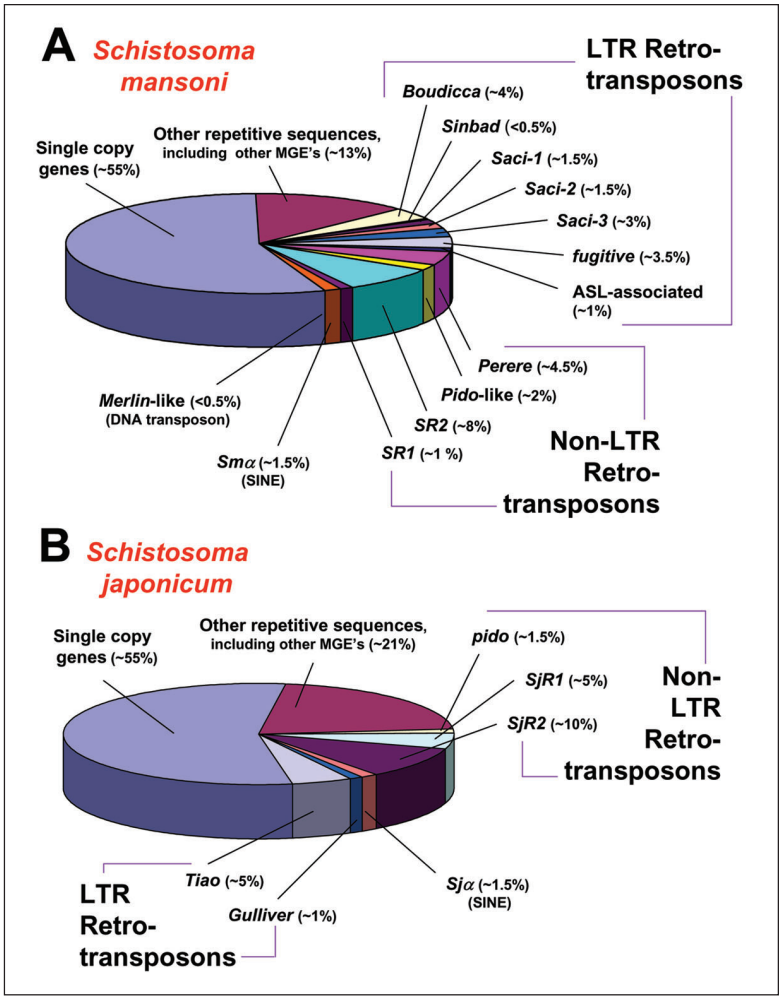

Figure 1. Contribution of mobile genetic elements in the genomes of (A) *Schistosoma mansoni* and (B) *Schistosoma japonicum* expressed in pie chart format. Predicted approximate contributions (in terms of simple DNA mass) of various mobile genetic elements, deducted from published and gene database sources are illustrated as pie slices, with percentages in text following the name of each mobile genetic element or genome category.

These elements are typically 5 - 10 kb in length. Their general structure consists of two open reading frames (ORFs) flanked by long direct terminal repeats of ~200-600 bp in length (Fig. 2). Some, such as *gypsy* from *Drosophila melanogaster*, *Osvaldo* from *Drosophila buzzatii*, and *Tas* from *Ascaris lumbricoides*, include a third ORF, *env*, encoding the envelope glycoprotein characteristic of retroviruses. Retroviruses probably evolved from LTR retrotransposons, mediated by the acquisition of envelope proteins that facilitated extracellular existence and horizontal transmission between cells and species.[19,20] The LTRs play a pivotal role in transposition and the initiation of transcription, with the 5' LTR encoding the promoter. The first ORF, *gag*, encodes a polyprotein precursor that is later processed to yield the structural proteins making up the virion core (Fig. 3). Of these, the nucleocapsid protein associates directly with the RNA, and

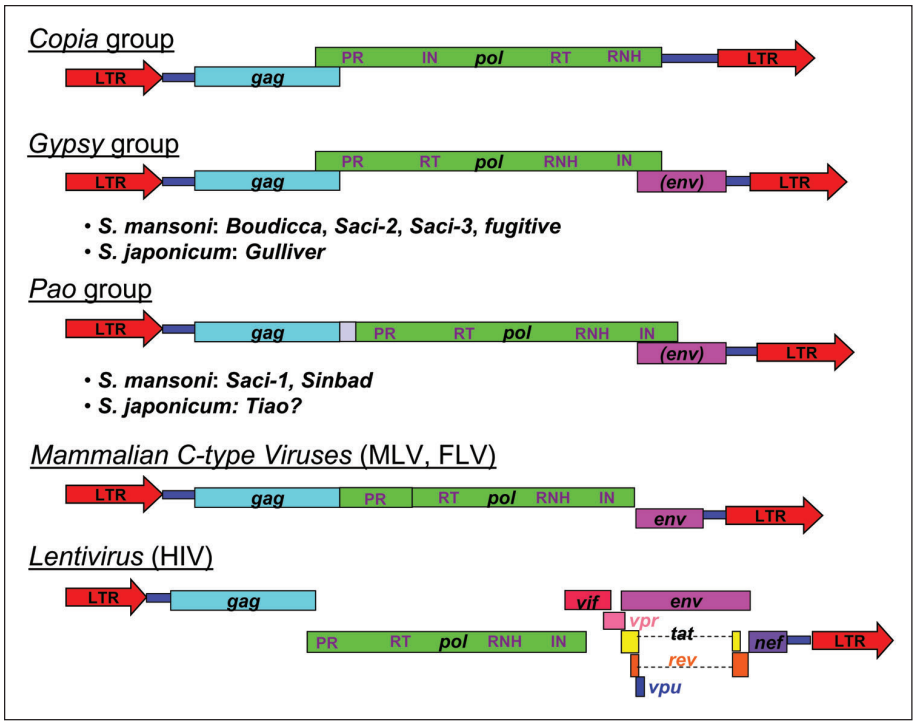

Figure 2. Diagrammatic representation of the genome structures of the three major LTR retrotransposon groups, *Copia/ Ty1*, *Gypsy/Ty3* and *Pao/BEL,* and two related retroviruses, the Mammalian C-type Viruses and the Lentiviruses. The LTR retrotransposons so far characterized from the *S. mansoni* and *S. japonicum* genomes fall into the *Gypsy/Ty3* and *Pao/BEL* groups, and are listed below the structures characteristic of these groups. Abbreviations: LTR, long terminal repeat; *gag*, group associated antigen; *pol*, polyprotein; env, envelope; PR, protease, RT, reverse transcriptase, RH, RNaseH, IN, integrase. The HIV genes *vif, vpr, rev, vpu,* and *nef* are accessory proteins specific to the more complex replication cycle of HIV.

exhibits a characteristic cysteine/histidine motif, which appears to function as a zinc finger domain. Transcription of the second ORF, *pol*, is initiated by read-through or ribosome slippage from the *gag* transcript. *Pol* encodes the enzymes of the retrotransposon, and, like *gag*, is cleaved by protease to yield functional proteins. These enzymes include protease, which cleaves the polyprotein, reverse transcriptase, which is responsible for transcribing the RNA genome into double stranded DNA, RNAse H, which removes the original RNA template from the newly synthesized DNA strand, allowing the complementary DNA strand to form, and integrase, which inserts the newly replicated proviral form of the retrotransposon into the host genome (Fig. 3). In retroviruses and LTR retrotransposons with an *env* gene, the envelope protein associates with the cell membrane, which envelops the virion core, allowing the viral particle to bud off from the host cell (Fig. 3). Envelope facilitates infection via attachment to specific cell surface receptors. Thus, in addition to vertical transmission in the germ line, LTR retrotransposons with *envelope* genes are capable of extracellular existence and horizontal transmission.

The two best understood retrotransposon families, *Copia/Ty1* and *gypsy/Ty3*, can be differentiated on the basis of their *pol* domain order, with a domain order of PR-RT-RNAse H-IN for the *gypsy/Ty3* retrotransposons and PR-IN-RT-RNAse H for the *Copia/Ty1* retrotransposons (Fig. 2). In addition to the *gypsy/Ty3* and *Copia/Ty1* families, two other families of LTR-retrotransposons are present in the schistosome genome that do not fit into either

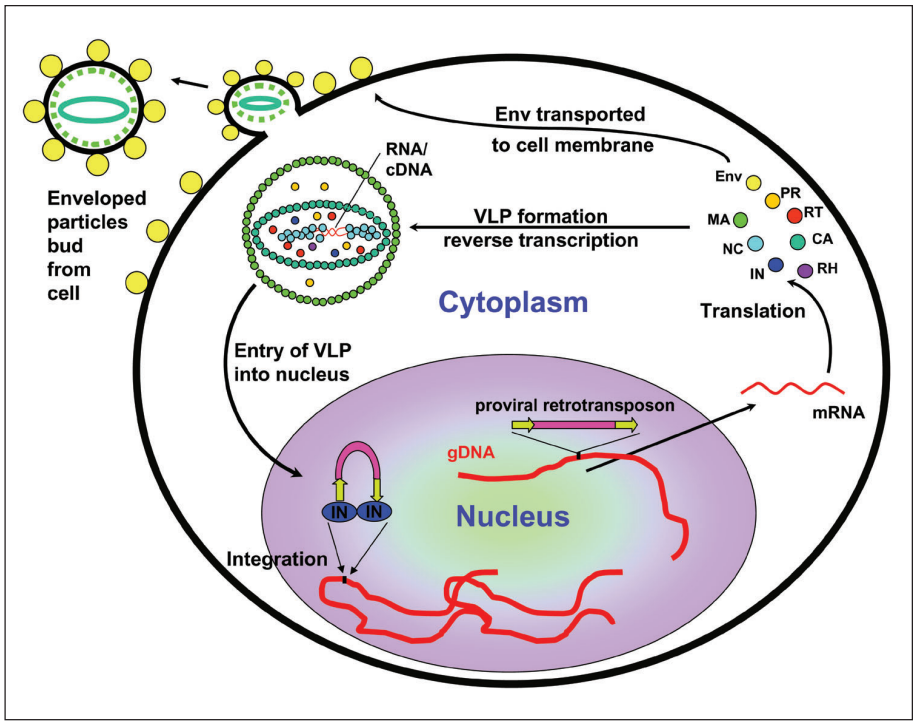

Figure 3. Diagrammatic representation of the replication cycle of an LTR retrotransposon. The proteins making up the virion core, NC, CA, and MA, are encoded by *gag*, whereas the enzymes PR, RT, RH, and IN are encoded by *pol*. *Env*, if present, is produced by splicing and is not a part of *gag* or *pol*. The VLP of LTR-retrotransposons is analogous to the virion of retroviruses. Abbreviations: gDNA, genomic DNA; mRNA, messenger RNA; cDNA, complementary DNA; NC; nucleocapsid; CA, capsid; MA, matrix; env, envelope; PR, protease, RT, reverse transcriptase, RH, RNaseH, IN, integrase; VLP, Viral-Like Particle. The illustration was adapted from Havecker et al[67] (with permission), using structural information from Coffin et al.[68]

of these clades. These include the *Pao* group retrotransposons, which have the same structural domain order as the *gypsy/Ty3* family, but form a distinct group at the primary sequence level, and the *Penelope* group of retrotransposons. The long terminal repeats of the *Penelope* retrotransposons, known as PLTRs, are distinctive in structure from the LTRs of LTR-retrotransposons, and indeed the *Penelope* elements were initially designated as a novel branch of nonLTR retroelements.[21] The open reading frames of the element extend into each PLTR, and the 5' PLTR is often preceded by an inverted PLTR.[21,22] The presence of introns in some members of this group, indicating the ability to replicate using an RNA-independent pathway, along with several other unique features indicate that these elements are distinct from both LTR and nonLTR retrotransposons.

Contributions of Transposable Elements to Genome Size, Structure and Evolution

The comparatively large size of the genome of *Schistosoma* is to a large extent due to the presence of numerous mobile genetic elements (Fig. 1). In addition to their contribution to genome size, retrotransposons also serve as a source of genetic variation by numerous means.[23] Retrotransposons appear to be quite active in schistosomes, with transposable element ESTs

making up from 3% (adult worms) to 14% (cercariae) of the transcriptome in *S. mansoni.*[24] The movement of retrotransposons can lead to insertional mutagenesis if they integrate into the coding portions of genes (see, for example, Casacuberta and Santiago).[25] Insertion into noncoding regions can also bring about significant changes. Integration into promoters or other regulatory elements can silence or enhance gene expression through disruption of these regulatory elements, and insertion into introns can lead to alternative splicing, resulting in exon skipping and truncated proteins.[26,27] Also, the LTRs themselves include transcriptional promoters, regulators, and terminators that can influence the transcription of a gene if the retrotransposon integrates near the gene.[25] Another epigenetic phenomenon that can come into play when transposable elements insert near genes is the spread of gene silencing from the element to the nearby gene. This type of gene silencing due to transposon-induced RNA interference (RNAi) has been observed in *Drosophila.*[28] Interestingly, transposable element activity seems to increase in the life stages of *S. mansoni* related to entry of the human host: transposable elements make up 14% and 7.4% of ESTs of cercariae and schistosomula, respectively, whereas they make up 3.0%, 3.7%, 4.9%, and 3.0% of ESTs of adults, eggs, miracidia, and mother sporocysts (germballs), respectively.[24] It is intriguing to speculate that this increased activity in cercariae and schistosomula may have evolved in response to the advantage conferred on a schistosome population by an increase in genetic diversity at the point of interface with the hostile immune system of the human host.

Genes can also be derived from transposable elements. The human genome contains dozens of genes that appear to be derived from transposable elements,[29] and *Arabidopsis* has been shown to have derived both promoter elements and exons from transposable elements.[30]

Finally, multi-copy elements can influence genomic structure through homologous recombination, resulting in chromosomal rearrangements such as deletions, duplications, or reciprocal translocations.[31,32] One example of a large scale, unexplained deletion in the schistosome genome is found within the cathepsin D gene of *S. mansoni*. The *S. mansoni* cathepsin D gene is homologous at the primary sequence level and has a similar exon-intron boundary structure to that the gene encoding cathepsin D in vertebrates, with one notable exception: two large exons appear to have undergone complete deletion.[33] The introns of this schistosome gene are large and highly populated by retrotransposons and retrotransposon fragments, which suggested to Morales et al[33] that homologous recombination between retrotransposon copies in two introns may have facilitated deletion of the exon between them.

Transposable elements, then, are capable of inducing mutagenesis and, as a result, variability in a population. Indeed, it has been suggested by Barbara McClintock and others that genomes have evolved to activate transposable elements in response to environmental stress.[34] Whether or not genomes have evolved mechanisms to specifically harness the power of transposable element induced mutation, with such a range of influence on genome change and variability, transposable elements clearly play a major role in genome evolution.

Schistosome Retrotransposable Elements: Gypsy/Ty3 Retrotransposons

Boudicca

Boudicca is a ~6 kb long terminal repeat retrotransposon from *S. mansoni*[35] (5,858 bp long for the original, well characterized 53-J-5 copy of the retrotransposon obtained from a bacterial artificial chromosome [BAC] library of the *S. mansoni* genome.[36] Two 328bp LTRs flank a coding region consisting of open reading frames representing the *gag* and *pol* polyproteins, 5' and 3' untranslated regions, and, at least in the BAC 53-J-5 copy, an additional unknown third open reading frame (Fig. 2). *Boudicca* is a high copy number element, estimated at 2,000 to 3,000 copies per haploid genome, and is actively transcribed in adult worms, cercariae, and sporocysts. One of several characterized members of the *Kabuki/CsRn1* clade of *gypsy*-type LTR retrotransposons, *Boudicca*'s closest nonschistosome relatives are *PwRn1* from the lungworm

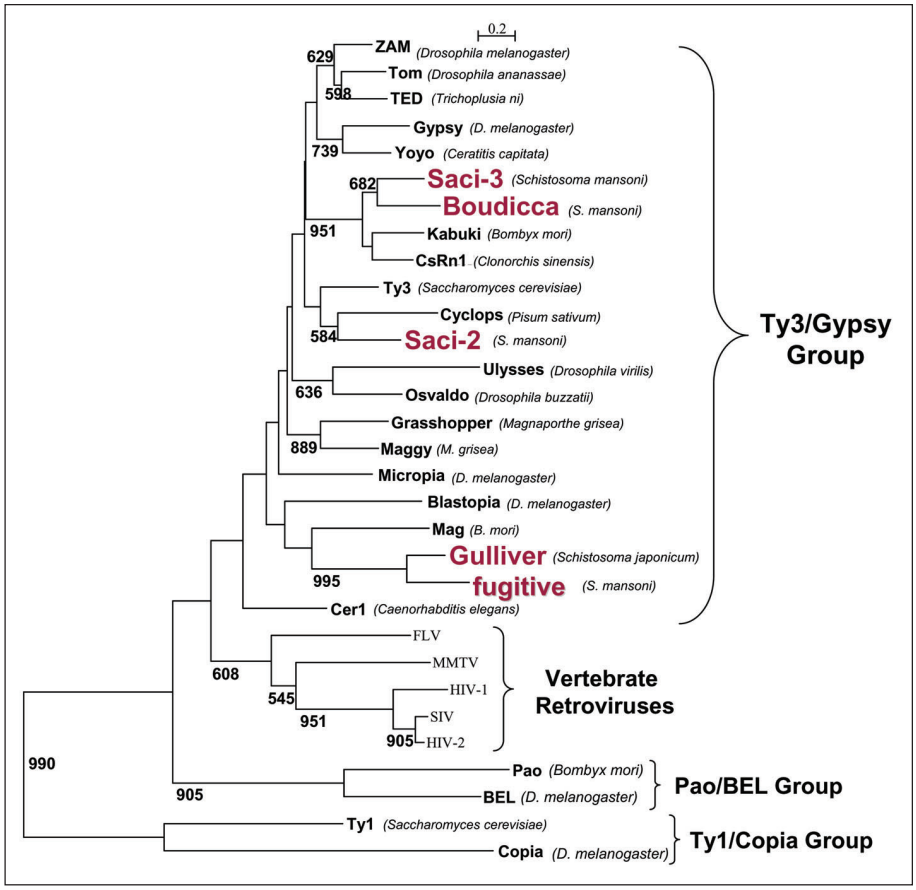

Figure 4. Phylogeny of *Gulliver, Boudicca, Saci-2, Saci-3*, the *fugitive,* and related elements. Phylogenetic tree comparing the relationships among the schistosome *gypsy* group LTR retrotransposons *Gulliver*, *Boudicca*, *Saci-2*, *Saci-3*, the *fugitive* and other LTR retrotransposons and retroviruses. Reverse transcriptase regions were aligned and tree files were generated using MacVector (Accelrys) and ClustalX software,[69] and trees were constructed using njplot software.[70] Gaps were excluded for the analysis, and a Clustal distance correction was employed to correct for multiple substitutions. Accession numbers: *BEL*: U23420, *Blastopia*: CAA81643, *Boudicca*: AY662653, *Cer1*: AAA50456, *Copia*: OFFFCP, *CsRn1*: AAK07486, *Cyclops*: AAL06415, HIV-1: P04585, HIV-2: AAA76841, FLV: NP_047255, *Fugitive*: BK005226, *Grasshopper*: AAA21442, *Gulliver*: AF243513, *Gypsy*: GNFFG1, *Kabuki*: BAA92689, *Mag*: S08405, *Maggy*: AAA33420, *Micropia*: S02021, MMTV: GNMVMM, *Osvaldo*: CAB39733, *Pao*: composite constructed from S33901, AB042118, and AB042119 (see Copeland et al.42), *Saci-2*: BK004069, *Saci-3*: BK004070, SIV: AAA47606, *TED*: AAA92249, *Tom*: CAA80824, *Ty1*: P47100, *Ty3*: AAA35184, *Ulysses*: CAA39967, *Yoyo*: T43046, *Zam*: CAA04050.

Paragonimus westermani, *CsRn1* from the oriental liver fluke, *Clonorchis sinensis*, and *Kabuki* from the silkworm, *Bombyx mori*. This clade, closely related to the errantiviruses, is differentiated not only at the primary sequence level of reverse transcriptase (Fig. 4) but also by a unique Cys-His box structure, CHCC instead of the more common retroviral CCHC Cys-His box.[37] Models of the secondary structure of the *Boudicca* transcript suggested that the first AUG was occluded by a stem loop structure, which in turn suggested a method of regulation of expression, at the level of translation, of *Boudicca* proteins.[38]

Fugitive

The *fugitive* element of *S. mansoni* belongs to the *mag*-like family of the *gypsy-Ty3* clade of long terminal repeat (LTR) retrotransposons.[39] It is closely related to the *mag*-like retrotransposon *Gulliver* from *S. japonicum*, but is dissimilar to several other LTR retrotransposons known from *S. mansoni* including *Boudicca*, *Saci-1*, *Saci-2*, *Saci-3*, and *Sinbad*. The full length *fugitive* element is 4,811 bp constituted of a single read-through open reading frame (ORF) of 4134 bp flanked at both ends by identical long terminal repeat sequences of 273 bp. This single ORF encodes retroviral-like *gag* and *pol* polyproteins, with a pol domain order of protease, reverse transcriptase, RNaseH and integrase. Gag of *fugitive* exhibits the distinctive double Cys-His motif characteristic of the *mag*-like retrotransposons.[40] Examination of schistosome transcriptome sequences in the public domain revealed that the *fugitive* was transcribed in at least six developmental stages of *S. mansoni*, while bioinformatics approaches and Southern hybridization analysis indicated that as many as 2000 copies of the *fugitive* were interspersed throughout the *S. mansoni* genome. Interestingly, the sequence of the first characterized copy (from BAC clone 49-J-14; accession no. BK005226) does not include any obvious mutations (stop or frame shift mutations). This apparent integrity, together with the identical sequences of the LTRs, suggests that the sequenced copy may be an active copy of this retrotransposon.

Gulliver

The other *mag*-family retrotransposon reported to date from schistosomes is *Gulliver*, from *Schistosoma japonicum*. *Gulliver* is ~4.8 kb in length and is present in multiple copies in the genome of *S. japonicum*[41] (Fig. 1). The LTRs of *Gulliver* are 259 bp in length and include RNA polymerase II promoter sequences, a CAAT signal and a TATA box. *Gulliver* exhibits features characteristic of a functional LTR retrotransposon including two read through (termination) ORFs encoding retroviral gag and pol proteins of 312 and 1,071 amino acid residues, respectively. The *gag* ORF encodes motifs conserved in nucleic acid binding proteins, while the *pol* ORF encodes conserved domains of aspartic protease, reverse transcriptase (RT), RNaseH and integrase, in that order, a pol pattern conserved in the *gypsy* lineage of LTR retrotransposons. *Gulliver*'s closest nonschistosome relatives include *mag* from *B. mori* and *Blastopia* from *D. melanogaster* (Fig. 4). Southern hybridization analysis indicated the presence of *Gulliver*-like elements in the genome of *S. mansoni,*[41] and these may be identical to the *fugitive* retrotransposon reported subsequently from *S. mansoni.*[39]

Saci-2

The *Saci-2* element of *S. mansoni* is 4,946 bp long and encodes a single ORF.[24] It is closely related to *Ty3* from *Saccharomyces cerevisiae*. *Saci-2* exhibits an unusual Cys-His box motif, with a lysine replacing a normally conserved histidine to yield a Cys-His box of the structure C(X2)C(X4)K(X4)C rather than the more conserved C(X2)C(X4)H(X4)C. This lysine replacement was conserved in 22 of 24 ESTs covering that region of the retrotransposon, and neither of the other two ESTs had a histidine in the ninth position, indicating that this replacement is not simply a mutation unique to a single *Saci-2* copy. *Saci-2* is transcribed in all six life stages of *S. mansoni.*[24]

Saci-3

Saci-3 is 5,217 bp long and is the second member of the *Kabuki/CsRn1* family to be found in the genome of *S. mansoni.*[24] *Saci-3* is closely related to *Boudicca* in primary sequence structure, and also has three open reading frames. This similarity, combined with low bootstrap values for separation of their RT amino acid sequences, call strict differentiation of *Boudicca* and *Saci-3* into question and raise the possibility that they are in fact distantly diverged copies of the same retrotransposon. However, the LTRs of these two retrotransposons are dissimilar as determined by phylogenetic analysis (bootstrap value for their separation is 96%), and their lengths are different (~210-290bp for *Saci-3* vs. ~330 bp for *Boudicca*),

supporting the classification of these elements as closely related but distinct. Like *Boudicca*, *Saci-3* also shares the CHCC Cys-His box structure unique to the *Kabuki/CsRn1* clade and is transcribed in all six life stages of *S. mansoni*.

Copia/Ty1 Retrotransposons

No *copia*-like retrotransposons have been described (yet) in schistosomes. However, preliminary sequence data from the *S. mansoni* genome sequencing project of The Institute for Genomic Research (TIGR) reveals three *S. mansoni* gDNA BAC clones exhibiting significant ($p < 1$) similarity to the complete DNA sequence of *Copia* from *Drosophila melanogaster*. As of February 2005, these BAC sequences, 48-C-10, 31-L-12, and 49-A-11, were still classified as "preliminary sequence data" and so were not yet available in the public databases, but they are available from the TIGR database, http://www.tigr.org/tdb/e2k1/sma1/gene.shtml.

Pao/BEL Retrotransposons

Saci-1

The first *Pao*-like element to be found in the genome of *S. mansoni*, *Saci-1* is 5980bp in length, with LTRs of 848bp in length. *Saci-1* has a single ORF encoding a polyprotein 1680 amino acids long, including domains for protease, reverse transcriptase, RNAse H, and integrase, as well as a triple Cys-His box.[24] Cys-His boxes exhibiting three separate zinc finger domains are unique to and characteristic of *Pao/BEL* elements. *Saci-1* was originally described from ESTs found in six life stages of *S. mansoni*, indicating that this element is actively transcribed throughout the *S. mansoni* life cycle.

Sinbad

Sinbad, the second *Pao*-like element to be found in the *S. mansoni* genome, is 6,287 bp long and is flanked at its 5'- and 3'-ends by identical LTRs of 386 bp.[42] Like *Saci-1*, *Sinbad* displays a triple Cys-His RNA binding motif characteristic of *gag* genes of *Pao/BEL*–like elements,[43,44] followed by the enzymatic domains of protease, reverse transcriptase (RT), RNAseH, and integrase, in that order. Though their vastly different LTR lengths confirm that *Saci-1* and *Sinbad* are different elements, they are nevertheless closely related, as determined by phylogenetic analysis. Together with an uncharacterized element from the *Danio rerio* genome, *Sinbad* and *Saci-1* appear to make up a separate subfamily of *Pao/BEL* elements (Fig. 5). Southern hybridization and bioinformatics analyses indicated that there were about 50 copies of *Sinbad* in the *S. mansoni* genome. The presence of ESTs spanning *Sinbad* in six life stages of *S. mansoni*, along with the identical 5'- and 3'-LTR sequences, indicate that *Sinbad* is actively transcribed throughout the *S. mansoni* life cycle.

Tiao

Tiao, the first LTR type retrotransposon to be reported from any of the human schistosomes, is a high copy number (about 10,000) LTR retrotransposon found in *S. japonicum* (GenBank AF073334).[45] *Tiao* is related to the *Pao*-like retrotransposons *Kamikaze, Yamato*, and *Pao* from *B. mori*, and *Ninja, BEL*, and *Max* from *Drosophila melanogaster*. More sequence data are needed to determine more precisely where *Tiao* falls within the phylogeny of *Pao/BEL* elements.

Other Schistosome LTR- retrotransposons

Cercyon

The *Penelope*-like retrotransposons, typified by *Penelope* of *Drosophila virilis* and *Xena* of *Takifugu rubripes* and *Tetraodon nigroviridis*, compose a highly divergent retrotransposon group distinct from both the LTR- and nonLTR retrotransposons. Though bounded by long terminal direct repeats, these elements lack Gag-, protease-, and RNAseH encoding regions, and

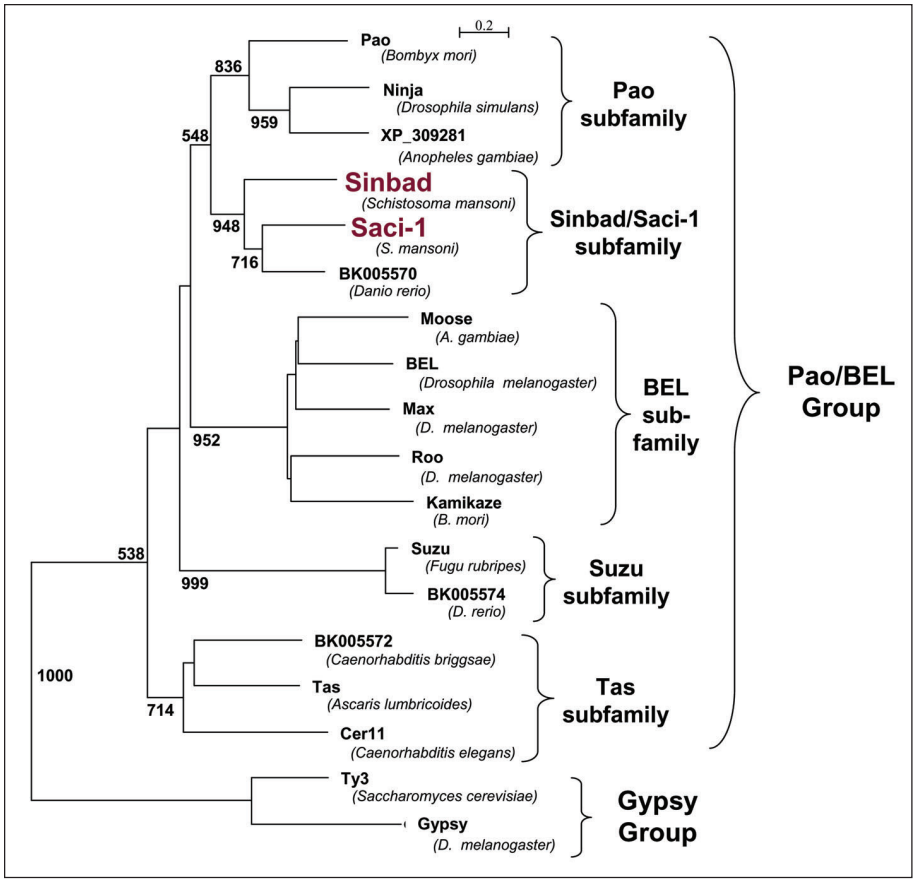

Figure 5. Phylogenetic tree comparing the relationships among the schistosome LTR retrotransposons *Sinbad, Saci-1* and other *Pao/BEL* retrotransposons, with *Ty3* and *Gypsy* as the outgroup. Reverse transcriptase regions were aligned and tree files were created using MacVector (Accelrys) and ClustalX software,[69] and trees were created using njplot software.[70] Gaps were excluded for the analysis, and a Clustal distance correction was employed to correct for multiple substitutions. Accession numbers: *BEL*: U23420, *Cer11*: AAA82437, *Gypsy*: GNFFG1, *Kamikaze*: AB042120, *Max*: CAD32253, *Moose*: AF060859, *Ninja*: T31674, *Pao*: composite constructed from S33901, AB042118, and AB042119 (see Copeland et al[42]), *Roo*: AAN87269, *Saci-1*: DAA04498, *Sinbad*: AY506538, *Suzu*: AF537216, *Tas*: Z29712, *Ty3*: AAA35184.

appear to replicate in a manner more similar to that of the nonLTR retrotransposons, as indicated by an abundance of 5' truncated copies and variable target site duplication lengths similar to those of L1-type elements.[22,46] Unique to the *Penelope*-like elements is the phenomenon of open reading frames extending into both the 5' and 3' long terminal repeats, leading to their designation as "PLTRs" to distinguish them from the LTRs of *Gypsy*-like, *Copia*-like, and *Pao*-like retrotransposons.[22] Indicative of a distinct mechanism of replication is the presence of an inverted PLTR flanking the 5' direct PLTR in many members of this group.[21,22] The *Penelope* elements are the only known retroelements to contain introns,[46] a further indication of a unique mechanism of replication. The *S. mansoni* genome contains a *Penelope*-like retrotransposon, designated *Cercyon* (GenBank accession no. BK000685). *Cercyon* is 2370 bp long, encodes reverse transcriptase and endonuclease domains, and is closely related to an unnamed element from the whipworm, *Trichuris*.[46]

ASL-Associated and Undescribed LTR Retrotransposons

Within the locus of the *S. mansoni* gene encoding adenylosuccinate lyase, intron 6 contains two 500 bp direct LTRs flanking a 1.9 kb sequence with homology to LTR retrotransposon type *gag* and *pol.*[47] Though probably truncated, this was the first retrotransposon to be identified in *S. mansoni*. In addition, at least three other LTR type retrotransposons related to *Pao*, *Mag*, and *Osvaldo* are present in the genome of *S. mansoni,* awaiting characterization (Copeland and Brindley, unpublished).

Applications to the Study of Schistosomes

Genome Sequencing Issues—Shotgun Sequencing

The advent of shotgun sequencing has served to accelerate the process of genome sequencing. The speed and efficiency of this method, however, comes at the cost of uncertainty at the point of joining individual runs to produce large contigs.[48] Multi-copy elements, and particularly high copy number elements such as many LTR retrotransposons, pose a significant problem at the stage of construction of larger contigs from individual sequence runs. Multi-copy elements raise the spectre of contigs including one copy of the element being falsely joined to contigs including a different copy of the element, or a different element with a similar sequence, as with reverse transcriptase, which includes well conserved domains.[49] Detailed understanding of these elements, including their sequences, their inter-copy variability, and their copy number, can serve to alert analysts to "high risk" repeat regions and the need for further scrutiny in such regions, using for example methods such as those detailed by Tammi et al,[48] to ensure the quality of sequencing of the *Schistosoma* genomes by this method.

Development of Transgenesis/Mutagenesis Vectors

The ability to introduce transgenes into genomes of pathogenic organisms has revolutionized molecular approaches to microbial diseases. Whereas this revolution has included protozoan parasites, and indeed mosquito and other insect vectors, so far it has largely bypassed medical helminthology because neither cell lines nor transgenesis systems have been developed for parasitic worms. Consequently, molecular research in schistosomes and schistosomiasis, and indeed other helminthoses, has been structurally disadvantaged in comparison with other microbes.[50] Since MGEs are valued as vectors for genetic transformation of other invertebrate genomes,[51] likewise they hold promise for genetic transformation of schistosomes. The use of transposons, such as the *Merlin*-like transposon[52] of *S. mansoni*, holds potential as a transformation strategy, as does the use of retroviruses, which have demonstrated utility in some insect species.[53] Though vertebrate retroviruses such as Moloney murine leukemia virus (Mo-MLV) have been widely used for transgenesis studies, including human gene therapy research, a chief drawback with retroviruses is that they typically exhibit narrow host ranges. Though this can be mitigated by procedures such as pseudotyping, an ideal LTR-type transgenesis vector would be one that already integrates into the schistosome genome as part of the vector's natural life cycle. The active search for and characterization of endogenous LTR retrotransposons in schistosomes have already yielded several possible candidates (see Figs. 4, 5). It seems feasible that schistosome retrotransposons such as these can be adapted directly as transgenesis vectors given similar success with other mobile elements.[54-56]

Insertional Mutagenesis

The goal of transgenesis leads to many other secondary objectives. One of these is the generation of collections of mutants, to enable examination of gene function through reverse genetics. Mobile genetic elements can be used to accomplish this goal. Transposable elements have been used for large scale insertional mutagenesis for the purposes of creating mutant assemblages of *Arabidopsis thaliana,*[57] rice,[58] *Drosophila melanogaster,*[59] *C. elegans*[60] and the zebrafish, *Danio rerio.*[61] The use of a marker gene flanked by LTRs of an endogenous schistosome LTR retrotransposon would be one of many ways to use these elements to deduce gene

function through reverse genetics studies. An increased understanding of insertion sites and specificity/generality of individual schistosome LTR retrotransposons would facilitate the foundation for these kinds of studies of the schistosome genome.

Knockout of Cell-Cycle Regulatory Genes to Create a Cell Culture Line

Schistosome research has been inhibited by the lack of a cell culture line for any of the schistosome species. Stable cell culture lines from humans and other species have been created from cancer cells, which have themselves been naturally transformed through destruction of protective cell-cycle regulatory genes, such as p53 (e.g., ref. 62). If analogous cell-cycle regulatory genes were to be discovered in schistosomes, their disruption could lead efforts to establish stable schistosome cell lines. One approach to such disruption could be the use of retrotransposons for insertional mutagenesis.

Other Applications

In addition, retrotransposons and other MGEs can be of value in comparative genomics. For example, the difference in size between the *Caenorhabditis briggsae* (~104 MB) and *C. elegans* (100.3 MB) genomes is almost entirely due to repetitive sequence, which accounts for 22.4 % of the C. briggsae genome in contrast to 16.5 % of the *C. elegans* genome. Few, if any, repeat families are shared, suggesting that most were acquired after the two species diverged or are undergoing rapid evolution.[63] Further, MGEs could provide paleontological information about the evolution of the genes at their integration sites, for example, in relation to orthologous target genes in sister host taxa.[64,65] Due to their interspersed nature in genomes, and their ability to move within genomes, PCR based techniques targeting these kinds of elements can be employed for identification of insertion/deletion polymorphic markers, and/or of transgene integrations (e.g., ref. 66). Finally, MGEs should also be of value in epidemiological and clinical studies because the interspersed nature of these elements can produce genetic fingerprints, which might be associated with clinically or otherwise relevant traits.

Acknowledgements

This work was supported in part by a Tulane Graduate School dissertation fellowship to CSC and a Burroughs Wellcome Fund Scholar Award in Molecular Parasitology to PJB.

References

1. Crompton DWT. How much human helminthiasis is there in the world? J Parasitol 1999; 85:397-403.
2. Chitsulo L, Loverde P, Engels D. Schistosomiasis. Nature Reviews Microbiology 2004; 2:12-13.
3. Lambertucci JR. Schistosoma mansoni: Pathological and clinical aspects. In: Jordan P, Webbe G, Sturrock RF, eds. Human Schistosomiasis. Wallingford, UK: CAB International, 1993:195-235.
4. Nokes C, McGarvey ST, Shiue L et al. Evidence for an improvement in cognitive function following treatment of Schistosoma japonicum infection in Chinese primary schoolchildren. Am J Trop Med Hyg 1999; 60:556-65.
5. Poggensee G, Feldmeier H. Female genital schistosomiasis: Facts and hypotheses. Acta Trop 2001; 79(3):193-210.
6. Gad A, Tanaka E, Orii K et al. Relationship between hepatitis C virus infection and schistosomal liver disease: Not simply an additive effect. J Gastroenterol 2001; 36(11):753-8.
7. Haitel A, Posch B, El-Baz M et al. Bilharzial related, organ confined, muscle invasive bladder cancer: Prognostic value of apoptosis markers, proliferation markers, p53, E-cadherin, epidermal growth factor receptor and c-erbB-2. J Urol 2001; 165(5):1481-7.
8. Qiu DC, Hubbard AE, Zhong B et al. A matched, case-control study of the association between Schistosoma japonicum and liver and colon cancers, in rural China. Ann Trop Med Parasitol 2005; 99(1):47-52.
9. Bergquist N. Schistosomiasis: From risk assessment to control. Trends Parasitol 2002; 18:309-314.
10. Chitsulo L, Engels D, Montresor A et al. The global status of schistosomiasis and its control. Acta Trop 2000; 77:41-51.
11. Zhou X, Chen M, McManus D et al. Schistosomiasis control in the 21st century. Proceedings of the International Symposium on Schistosomiasis, Shanghai, July 4-6, 2001. Acta Tropica 2002; 82:95-114.

12. Markell EK, John DT, Krotoski WA. Markell and voge's medical parasitology. 8th ed. W.B. Saunders Company, a division of Harcourt Brace & Company, 1999.
13. Simpson AJG, Sher A, McCutchan TF. The genome of Schistosoma mansoni: Isolation of DNA, its size, bases and repetitive sequences. Mol Biochem Parasitol 1982; 6:125-137.
14. Hirai H, Taguchi T, Saitoh Y et al. Chromosomal differentiation of the Schistosoma japonicum complex. Int J Parasitol 2000; 30(4):441-452.
15. Verjovski-Almeida S, DeMarco R, Martins EA et al. Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni. Nat Genet 2003; 35:148-157.
16. Hu W, Yan Q, Shen D-K et al. Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource. Nat Genet 2003; 35:139-147.
17. Johnston DA, Blaxter ML, Degrave WM et al. Genomics and the biology of parasites. Bioessays 1999; 21(2):131-47.
18. Brindley PJ, Laha T, McManus DP et al. Mobile genetic elements colonizing the genomes of metazoan parasites. Trends Parasitol 2003; 19:79-87.
19. Temin HM. Origin of retroviruses from cellular moveable genetic elements. Cell 1980; 21:599-600.
20. Malik HS, Henikoff S, Eickbush TH. Poised for contagion: Evolutionary origins of the infectious abilities of invertebarate retroviruses. Genome Res 2000; 10:1307-1318.
21. Lyozin GT, Makarova KS, Velikodvorskaja VV et al. The structure and evolution of Penelope in the virilis species group of Drosophila: An ancient lineage of retroelements. J Mol Evol 2001; 52(5):445-56.
22. Dalle Nogare DE, Clark MS, Egar G et al. Xena, a full-length basal retroelement from tetraodontid fish. Mol Biol Evol 2002; 19:247-255.
23. Kazazian Jr HH. Mobile elements: Drivers of genome evolution. Science 2004; 303:1626-1632.
24. DeMarco R, Kowaltowski AT, Machado AA et al. Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite Schistosoma mansoni. J Virol 2004; 78(6):2967-78.
25. Casacuberta JM, Santiago N. Plant LTR-retrotransposons and MITEs: Control of Transposition and Impact on the Evolution of Plant genes and genomes. Gene 2003; 311:1-11.
26. Varagona MJ, Purugganan M, Wessler SR. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. Plant Cell 1992; 4(7):811-20.
27. Goldin SN, Papaioannou VE. Unusual misregulation of RNA splicing caused by insertion of a transposable element into the T (Brachyury) locus. BMC Genomics 2003; 4(1):14.
28. Puig M, Caceres M, Ruiz A. Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. Proc Natl Acad Sci USA 2004; 101(24):9013-8.
29. International genome sequencing consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409(6822):860-921.
30. Lippman Z, Gendrel AV, Black M et al. Role of transposable elements in heterochromatin and epigenetic control. Nature 2004; 430(6998):471-6.
31. Evgen'ev MB, Zelentsova, Helena et al. Mobile elements and chromosomal evolution in the virilis group of Drosophila. Proc Natl Acad Sci USA 2000; 97(21):11337-11342.
32. Lim JK, Simmons MJ. Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. Bioessays 1994; 16(4):269-75.
33. Morales ME, Kalinna BH, Heyers O et al. Genomic organization of the Schistosoma mansoni aspartic protease gene, a platyhelminth orthologue of mammalian lysosomal cathepsin D. Gene 2004; 338(1):99-109.
34. Lonnig WE, Saedler H. Chromosome rearrangements and transposable elements. Annu Rev Genet 2002; 36:389-410.
35. Copeland CS, Brindley PJ, Heyers O et al. Boudicca, a retrovirus-like, LTR retrotransposon from the genome of the human blood fluke, Schistosoma mansoni. J Virol 2003; 77:6153-6166.
36. Le Paslier MC, Pierce RJ, Merlin F et al. Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 2000; 65(2):87-94.
37. Bae YA, Moon SY, Kong Y et al. CsRn1, a novel active retrotransposon in a parasitic trematode, Clonorchis sinensis, discloses a new phylogenetic clade of Ty3/gypsy-like LTR retrotransposons. Mol Biol Evol 2001; 18(8):1474-1483.
38. Copeland CS, Heyers O, Kalinna BH et al. Structural and evolutionary analysis of the transcribed sequence of Boudicca, a Schistosoma mansoni retrotransposon. Gene 2004; 329:103-114.
39. Laha T, Loukas A, Smyth DJ et al. The fugitive LTR retrotransposon from the genome of the human blood fluke, Schistosoma mansoni. International Journal for Parasitology 2004; 34:1365-1375.
40. Michaille JJ, Mathavan S, Gaillard J et al. The complete sequence of mag, a new retrotransposon in Bombyx mori. Nucleic Acids Res 1990; 18(3):674.
41. Laha T, Loukas A, Verity CK et al. Gulliver, a long terminal repeat retrotransposon from the genome of the oriental blood fluke Schistosoma japonicum. Gene 2001; 264(1):59-68.

42. Copeland CS, Mann VH, Morales ME et al. The Sinbad retrotransposon from the genome of the human blood fluke, Schistosoma mansoni, and the distribution of related Pao-like elements. BMC Evol Biol 2005; 5(1):20.
43. Xiong Y, Burke WD, Eickbush TH. Pao, a highly divergent retrotransposable element from Bombyx mori containing long terminal repeats with tandem copies of the putative R region. Nucleic Acids Res 1993; 21:2117-2123.
44. Frame IG, Cutfield JF, Poulter TM. New BEL-like LTR-retrotransposons in Fugu rubripes, Caenorhabditis elegans, and Drosophila melanogaster. Gene 2001; 263:219-230.
45. Fan J, Brindley PJ. Retrotransposable elements in the Schistosoma japonicum genome. In: Isao Tada SK, Moriyasu Tsuji, eds. Ninth International Congress of Parasitology, (August 24-28, 1998). Makuhari Messe. Chiba, Japan: Monduzzi Editore, 1998:821-825.
46. Arkhipova IR, Pyatkov KI, Meselson M et al. Retroelements containing introns in diverse invertebrate taxa. Nat Genet 2003; 33(2):123-4.
47. Foulk BW, Pappas G, Hirai Y et al. Adenylosuccinate lyase of Schistosoma mansoni: Gene structure, mRNA expression, and analysis of the predicted peptide structure of a potential chemotherapeutic target. Int J Parasitol 2002; 32(12):1487-1495.
48. Tammi MT, Arner E, Kindlund E et al. Correcting errors in shotgun sequences. Nucleic Acids Res 2003; 31(15):4663-72.
49. Xiong Y, Eickbush TH. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 1988; 5(6):675-90.
50. Boyle JP, Yoshino TP. Gene manipulation in parasitic helminths. Int J Parasitol 2003; 33:1259-1268.
51. O'Brochta D, Sethuraman N, Wilson R et al. Gene vector and transposable element behavior in mosquitoes. J Exp Biol 2003; 206:3823-3834.
52. Feschotte C. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. Mol Biol Evol 2004.
53. Handler AM. A current perspective on insect gene transformation. Insect Biochem Mol Biol 2001; 31(2):111-28.
54. Garraway LA, Tosi LR, Wang Y et al. Insertional mutagenesis by a modified in vitro Ty1 transposition system. Gene 1997; 198:27-35.
55. Ivics Z, Hackett PB, Plasterk RH et al. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 1997; 91:501-510.
56. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 2003; 35:41-48.
57. Ostergaard L, Yanofsky MF. Establishing gene function by mutagenesis in Arabidopsis thaliana. Plant J 2004; 39(5):682-96.
58. Hirochika H. Contribution of the Tos17 retrotransposon to rice functional genomics. Current Opinion in Plant Biology 2001; 4:118-122.
59. Bellen HJ, Levis RW, Liao G et al. The BDGP gene disruption project: Single transposon insertions associated with 40% of Drosophila genes. Genetics 2004; 167:761-81.
60. Williams DC, Boulin T, Ruaud AF et al. Characterization of Mos1 mediated mutagenesis in C. elegans: A method for the rapid identification of mutated genes. Genetics 2005; Epub ahead of print.
61. Amsterdam A, Nissen RM, Sun Z et al. Identification of 315 genes essential for early zebrafish development. Proc Natl Acad Sci USA 2004; 101:12792-7.
62. Isaka K, Nishi H, Sagawa Y et al. Establishment of a new human cell line (EN) with TP53 mutation derived from endometrial carcinoma. Cancer Genet Cytogenet 2003; 141(1):20-5.
63. Stein LD, Bao Z, Blasiar D et al. The genome sequence of Caenorhabditis briggsae: A platform for comparative genomics. PLoS Biol 2003; 1:E45.
64. SanMiguel P, Gaut BS, Tikhonov A et al. The paleontology of intergene retrotransposons of maize. Nat Genet 1998; 20:43-45.
65. Batzer M, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002; 3:370-379.
66. Mamedov IZ, Arzumanyan ES, Amosova AL et al. Whole-genome experimental identification of insertion/deletion polymorphisms of interspersed repeats by a new general approach. Nucleic Acids Res 2005; 33(2):e16.
67. Havecker ER, Gao X, Voytas DF. The diversity of LTR retrotransposons. Genome Biol 2004; 5(6):225.
68. Coffin JM, Hughes SH, Varmus HE. Retroviruses. Plainview, NY: Cold Spring Harbor Laboratory Press, 1997.
69. Thompson JD, Gibson TJ, Plewniak F et al. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25:4876-4882.
70. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406-425.